Journal for Research in Mathematics Education
2008, Vol. 39, No. 3, 247–280

The Impact of Middle-Grades Mathematics Curricula and the Classroom Learning Environment on Student Achievement

James E. Tarr, Robert E. Reys, Barbara J. Reys, and Óscar Chávez
University of Missouri

Jeffrey Shih
University of Nevada—Las Vegas

Steven J. Osterlind
University of Missouri

We examine student achievement of 2533 students in 10 middle schools in relation to the implementation of textbooks developed with funding from the National Science Foundation (NSF) or publisher-developed textbooks. Using hierarchical linear modeling (HLM), curriculum type was not a significant predictor of student achievement on the Balanced Assessment in Mathematics (BAM) or TerraNova Survey (TNS) after controlling for student-level variables. However, the Standards-Based Learning Environment (SBLE) moderated the effect of curriculum type. Students were positively impacted on the *BAM* by NSF-funded curricula when coupled with either Moderate or High levels of SBLE. There was no statistically significant impact of NSF-funded curricula on students in classrooms with a Low level of SBLE, and the relationship between publisher-developed textbooks and SBLE was not statistically significant. Moreover, there was no significant impact of either curriculum type when coupled with varying levels of SBLE on the *TNS* as the dependent measure.

Key words: Achievement; Curriculum; Integrated Curriculum; Middle grades (5–8); Quasi-experimental design; Reform in mathematics education; Teaching practice

In this era of high-stakes testing and public accountability, school personnel are searching for ways to improve mathematics learning opportunities for all students. Although there is no single "magic bullet" solution, one avenue for strengthening

---

This study was funded by the United States Department of Education (Grant No. R303T010735), but this article does not reflect their opinion or endorsement. The first three authors were affiliated with the Show-Me Center when this research was conducted. The researchers gratefully acknowledge the willingness of schools to participate in this Middle School Mathematics Study. It required the approval of superintendents, principals, and in some cases mathematics coordinators. Most important, it required the willingness of classroom teachers and their students to participate. Without their participation, school-based research is not possible. Finally, we are thankful to Brian Townsend, David D. Barker, Jung-Chih Chen, Troy P. Regis, and Aina Appova for their many contributions to this project.

school mathematics programs is through the selection and implementation of high-quality mathematics textbooks. Researchers note that textbooks have historically played a prominent role in U.S. mathematics classrooms, often defining the mathematics curriculum that students have an opportunity to learn (Schmidt, McKnight, & Raizen, 1997; Tyson, 1997; Weiss, Banilower, McMahon, & Smith, 2001) and underscoring the need to give careful attention to what mathematics content is emphasized and how it is presented (Grouws & Smith, 2000; Peak, 1996). In fact, 89% of U.S. eighth graders report doing mathematics problems from their textbook (Lindquist, 1997) and nearly three fourths of eighth-grade teachers in the United States report using their textbook on a daily basis (Grouws & Smith, 2000). Two thirds of middle-grades mathematics teachers indicate that they use one mathematics textbook all or most of the time, and the same number indicate that they "cover" at least three fourths of the textbook in a given year (National Research Council [NRC], 2001a).

Despite the dominant role that mathematics textbooks have historically played in defining the school mathematics experience of students, drawing a direct link from the textbook to student learning is difficult because many other factors influence what students learn, including teacher choices and actions, school and classroom organization, and students' readiness and willingness to learn. Notwithstanding the complexity of factors that contribute to student learning, textbooks are viewed by many as an important lever for change—a tangible tool for impacting what teachers do and therefore what students learn. However, the content and quality of U.S. textbooks are dependent on commercial publishers that operate in a national, market-driven environment. Publishers produce textbooks that are sold nationally but must respond to local standards that vary from state to state (Reys, 2006). In some states, a list of "approved" textbooks is generated, and individual school districts make selections from this narrow list. In other states, school districts choose textbooks on their own timeline based on criteria developed locally. Thus, there is no central system that influences nationally developed textbooks (Reys & Reys, 2006; Sewall, 2005; Thomas B. Fordham Institute, 2004; Tyson-Bernstein & Woodward, 1991).

In response to the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989) and in an effort to influence and strengthen the quality of U.S. mathematics textbooks, the National Science Foundation (NSF) has invested an estimated $93 million in K–12 mathematics curriculum development efforts (NRC, 2004). Curriculum development teams consisting of mathematics educators, mathematicians, and classroom teachers worked together to produce mathematics textbooks that embodied "standards-based" characteristics, including active engagement of students, a focus on problem solving, and attention to connections within mathematical strands as well as to real-life contexts (Reys, Robinson, Sconiers, & Mark, 1999; Trafton, Reys, & Wasman, 2001).

The curriculum development efforts funded by NSF together with textbooks produced by commercial publishers offer school districts the largest variety of curricular materials available in the past 4 decades. Textbooks differ in their philosoph-

ical orientation (i.e., their view of student learning) as well as in content and in the nature and amount of material provided to teachers (Trafton, Reys, & Wasman, 2001). These differences have the potential to influence what mathematics teachers emphasize, how that content is presented, and most important, what mathematics students learn.

Researching the impact of mathematics textbooks on student learning is a complex endeavor (Sawada et al., 2002) that is difficult to conduct for many reasons. These reasons include gaining access to schools, documenting the extent to which teachers follow the textbook, collecting data over a sustained period of time, identifying appropriate comparison groups, isolating variables, and accessing valid measures of student achievement (Chval, Chávez, Reys, & Tarr, in press; Hiebert, 1999; Schoenfeld, 2000; Usiskin, 1998).

Notwithstanding the challenges inherent in studying the relationship among teachers, curriculum, and student learning, a growing body of studies provides evidence that NSF-funded curricular materials influence teacher decisions and actions (e.g., Remillard, 1999; Remillard & Bryans, 2004; Sawada et al., 2002) and positively affect student learning (e.g., Briars, 2001; Griffin, Evans, Timms, & Trowell, 2000; Huntley, Rasmussen, Villarubi, Sangtong, & Fey, 2000; Mullis et al., 2000; Reys, Reys, Lapan, Holliday, & Wasman, 2003; Riordan & Noyce, 2001). With respect to curricular effectiveness, the NRC (2004) identified 67 studies of NSF-funded curricula; 11 studies of University of Chicago School Mathematics Project (UCSMP); and 17 of commercially published curricula, 14 of which were conducted on Saxon materials. “As these results suggest, we know more about the evaluations of the NSF-funded curricula and UCSMP than about the evaluations of the commercial programs” (p. 99). The report further acknowledged that “there was some reported reluctance on the part of commercial companies to release studies that could affect perceptions of competitive advantage” (p. 100).

Variation in implementation of curricular materials is cited as a factor that is likely to contribute to mediation of the impact of textbooks on student learning (Briars, 2001; Remillard, 2005; Huntley et al., 2000; Schoenfeld, 2000; Tarr, Chávez, Reys, & Reys, 2006). Indeed, textbook implementation is an uneven process within and across schools (Grouws & Smith, 2000; Jackson, 1992; Senk & Thompson, 2003). As Kilpatrick (2003) argued:

> Two classrooms in which the same curriculum is supposedly being ‘implemented’ may look very different; the activities of teacher and students in each room may be quite dissimilar, with different learning opportunities available, different mathematical ideas under consideration, and different outcomes achieved. (p. 473)

Researchers are therefore urged to consider variations in implementation of curricular materials. The evaluation of mathematics curricula must be situated within this reality and certain minimum conditions must be established in order to examine links between textbooks and student learning outcomes. As noted by the NRC (2004),

> A standard for evaluation of any social program requires that an impact assessment is warranted only if two conditions are met: (1) the curricular program is clearly speci-

> fied, and (2) the intervention is well implemented. Absent this assurance, one must have a means of ensuring or measuring treatment integrity in order to make causal inferences. (p. 100)

The design of this study took into account "treatment integrity" by including teachers' use of available curricular materials and their provision of key instructional practices associated with *Standards*-based instruction.

CONCEPTUAL FRAMEWORK

The term "curriculum" is used in many different ways, both in practice and in the scholarly literature. For example, some researchers make a distinction between the *intended* curriculum (the official set of learning expectations as described in local, state, or national curriculum frameworks), the *implemented* curriculum (the mathematics presented to students by the teacher), and the *attained* curriculum (the mathematics that students learn) (Robitaille, 1992; Schmidt et al., 1997; Travers & Westbury, 1989). Porter and Smithson (2001) distinguished the intended from the *assessed* curriculum (represented by high-stakes tests), and the *enacted* (implemented) from the *learned* curriculum (the content that students learn).

Figure 1 illustrates how each interpretation of curriculum connects to and influences other components of the educational system. Note that we distinguish the intended curriculum from the "textbook" curriculum. That is, the textbook adopted by a school or district is often not perfectly aligned with the intended curriculum (the local or state mathematics curriculum framework); consequently, teachers must make decisions, often on a daily basis, about what to use from the textbook, what to skip, and what to supplement from other resources. In this sense, teachers are active developers of the enacted/implemented curriculum, influenced by experiences that occur within the mathematics classroom as well as by the instructional materials available to them (Ben-Peretz, 1990; Clandinin & Connelly, 1992; Clarke, Clarke, & Sullivan, 1996; Love & Pimm, 1996; Remillard, 1999; Remillard, 2005; Remillard & Bryans, 2004).

Figure 1 includes a brief description of various forces that influence each curriculum and, in turn, student opportunities to learn. For example, current educational policies such as No Child Left Behind (NCLB, 2002) call for increased articulation of learning goals accompanied by annual assessments to measure student progress in relation to these goals. Such policies and their resulting products (state or local curriculum standards documents and assessments) influence the textbook curriculum as well as the implemented curriculum (Jennings & Rentner, 2006). In addition, the teachers' own beliefs and experiences as well as the student's prior knowledge and motivation influence the ways in which the student interacts with mathematics. Instructional strategies, whether explicitly suggested by the textbook lesson or applied by the teacher independent of available curricular materials, are other key influences on student learning.

Within our conceptual framework, we focus primarily on links between the Textbook Curriculum, Implemented Curriculum, and Learned Curriculum. In

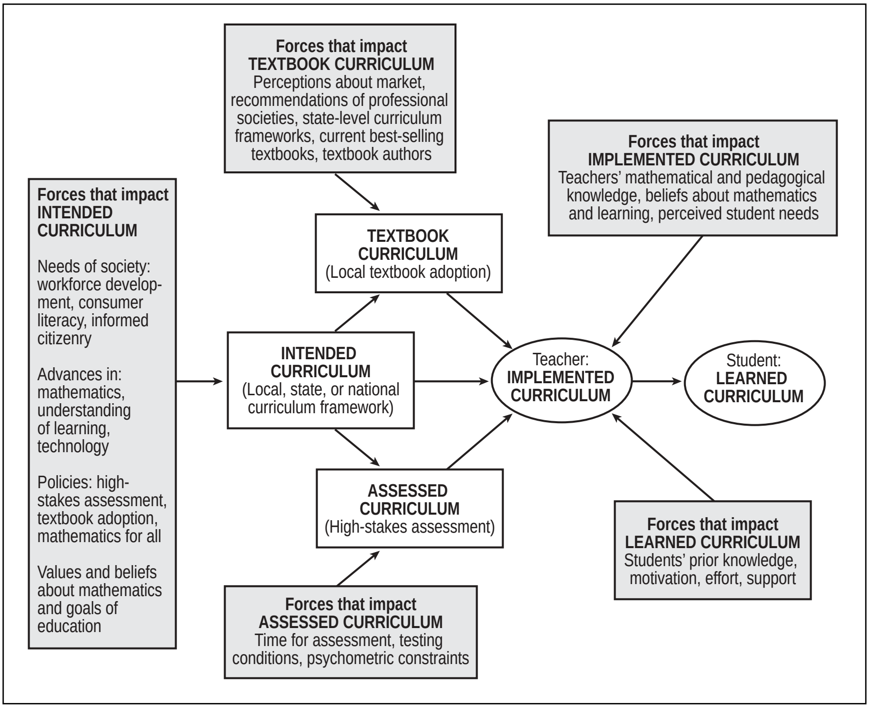

Figure 1. Model depicting the relationship of various types of curriculum and the forces that influence the content of those curricula.

particular, we examine the impact of two distinct types of Textbook Curriculum (NSF funded and publisher developed) on the Implemented Curriculum and, in turn, on the Learned Curriculum. Specifically, we examined relationships between the implementation of district-adopted textbooks, Standards-based Learning Environment (SBLE), and student achievement. In classrooms where textbooks were found to be a significant determinant of the students' opportunity to learn, analyses were undertaken. Consistent with our conceptual framework, we focused on three primary research questions:

1. Does the implementation of district-adopted textbooks differ by *curriculum type* (NSF funded vs. publisher developed)?
2. In classrooms where district-adopted textbooks are a strong determinant of the content and activity of mathematics lessons, to what extent does *curriculum type* (NSF funded vs. publisher developed) predict student achievement in mathematics?
3. Does the relationship between *curriculum type* and student achievement in mathematics differ based on variations in the SBLE?

METHODOLOGY

The research questions required a project design that captured the complexities of curriculum implementation and was simultaneously sensitive to the reality of school settings. Therefore, we used a quasi-experimental design and employed both qualitative and quantitative methods. Qualitative methods were used to characterize curriculum implementation by teachers, including the nature and extent of use of district-adopted textbooks. Moreover, consistent with our conceptual framework, we made classroom observations to document the extent to which teachers created a learning environment consistent with the tenets of the NCTM *Standards*.[1] Hierarchical linear modeling (HLM) was employed to ascertain relationships among curriculum, teaching, and learning outcomes. Collectively, these methods provided a database to profile teacher and student experiences with district-adopted textbooks over a 2-year period in order to understand the impact of middle-grades mathematics curricula on student achievement.

Identification of Sample

Schools. Three criteria were established to inform the selection of schools: (a) number of years since adoption of the current mathematics textbook series, (b) organizational structure of the middle school, and (c) size of the community in which the school was located. In particular, we did not consider schools in their 1st year of implementation of *any* textbook because research suggests that many schools gradually introduce new curricular materials before full implementation (Bay-Williams, Reys, & Reys, 2003; Senk & Thompson, 2003). Given the longitudinal nature of our study, preference was given to middle schools using a Grade 6–8 organizational structure. In avoiding use of elementary schools, we eliminated many problems associated with students changing schools and also consolidated communication with school administrators and teachers. Finally, we sought representation of a variety of community types to reflect the diversity of the U.S. school population. Hence, we solicited participation of rural, small community, suburban, and urban schools. These community types were operationally defined as follows: rural—strongly rural characteristics with a population less than 5,000; small city—population of 5,000 to 10,000 and not near or part of a larger population area; suburban—population of 5,000 to 100,000 near or part of a larger populated area; and urban—population over 100,000. We also sought diversity with respect to ethnic representation as well as a range of the percent of students qualifying for Free or Reduced Lunch (FRL).

Given the absence of a database of textbooks used in our nation's more than 15,000 school districts, we solicited Principal Investigators of NSF Instructional Materials Development (IMD) grants to provide names of school districts that were beyond the 1st year of curriculum implementation. From the list, we contacted

[1] Throughout the article "NCTM *Standards*" refers to a collection of four documents (NCTM, 1989, 1991, 1995, 2000).

school administrators, outlining goals of the study and responsibilities of school and research personnel. Once six schools using the NSF-funded textbooks were selected, we sought to identify and obtain permission from schools with similar demographics—particularly the percentage of students qualifying for FRL—that were using publisher-developed mathematics textbooks with a significant market share based on the 2000 Mathematics and Science Education Survey (Weiss et al., 2001).

Once schools were selected, we sought the permission of all students in Grades 6, 7, and 8 and their mathematics teachers to participate in the study. In addition to the promise of useful data yielded from the study, we also offered modest incentives to participate, including a $495 compensation for teachers' time (e.g., attending an orientation meeting, completing surveys, allowing classroom observations, administering and returning project-related exams) and a $250 school stipend for provision of student achievement data, class rosters, and ongoing communication with project staff, including local data collectors. Despite these incentives, only approximately 1 in 3 districts ultimately agreed to participate. It should be noted that reluctance to participate was evident in both types of schools, regardless of whether the district-adopted textbook was publisher developed or NSF funded. (For a full description of the challenges in recruiting schools, see Chval, Reys, Reys, Tarr, & Chávez, 2006.) Notwithstanding these challenges, across the 11 middle schools that accepted the invitation to participate, only 1 teacher (out of more than 70) and about 20 students (out of more than 4,000) declined to participate in both years of the study.

Although 11 schools agreed to participate in the study, student achievement data from only 10 schools were ultimately analyzed. One urban school using *Connected Mathematics Project* (Lappan, Fey, Fitzgerald, Friel, & Phillips, 1998) was dropped after it did not fulfill the commitment to provide prior achievement data, including FRL data at the student level. The decision to drop this school was further supported by the fact that the district experienced high mobility and student absenteeism in both years; in fact, more than half of all students were absent for at least one of the end-of-year tests. Moreover, one seventh-grade teacher (who taught more than half of the sixth graders in the previous year) declined to participate, citing an exhaustive list of prior commitments. Collectively these issues yielded a profoundly reduced sample of students from this school. In short, complete achievement data were not available for any student in this school, precluding any meaningful data analysis (implementation data, however, was not compromised). Despite the loss of 1 urban school, 2 urban schools remained in the sample for the duration of the study, ensuring representation of a diversity of student populations.

Teachers. In order to describe the educational background, beliefs, and instructional practices of participating teachers in the study, we amended a survey designed by Horizon Research, Inc. (Weiss et al., 2001) to solicit responses from teachers about their use of curricular materials to plan and enact instruction. We examined a number of attributes, including teacher beliefs, years of teaching experience, initial certification area, highest degree held, number of college-level mathematics courses, and amount and perceived impact of recent professional development.

A comparison of the two sets of teachers (those using publisher-developed curricula and those using NSF-funded curricula) indicates there were no significant differences in years of teaching experience, degrees held, or initial certification area. With respect to mathematics background, the two sets of teachers were generally comparable. In particular, although there were no significant differences in 12 of 14 surveyed content areas, teachers of publisher-developed curricula were more likely to report college coursework in abstract algebra, whereas teachers of NSF-funded curricula were more likely to report coursework in college algebra/trigonometry/elementary functions.

The two sets of teachers reported comparable familiarity with the NCTM *Standards,* indicated similar levels of agreement with the overall vision of mathematics education described in these documents, and expressed similar extent of implementation of NCTM's recommendations. With respect to professional development, teachers of NSF-funded curricula reported more time spent attending workshops on mathematics teaching or recent coursework in mathematics than teachers of publisher-developed curricula. These same teachers indicated professional development focused on learning how to use the textbook adopted by their district, with particular emphasis on learning to use inquiry/investigation-oriented methods. In this aspect, there were no "matching" professional development opportunities for teachers using publisher-developed textbooks.

With respect to teacher beliefs, in relation to teachers of NSF-funded curricula, teachers of publisher-developed curricula were more likely to believe that students learn best in classes with students of similar abilities, and students of these teachers were more likely to be enrolled in mathematics classes in which students were grouped by ability. Additionally, teachers of publisher-developed curricula were more likely to believe that students learn mathematics best when the teacher demonstrates concepts and methods and then provides students opportunities for practice and reinforcement. Teachers of NSF-funded curricula were more likely to report feeling prepared to encourage participation of minorities in mathematics. Finally, teachers of publisher-developed curricula were more likely to report that students worked individually, whereas teachers of NSF-funded curricular reported students tended to work collaboratively, in small groups.

In summary, the two groups of teachers were comparable in educational background, familiarity with national professional standards, and teaching licensure and experience, but they differed in the amount and perceived impact of recent professional development as well as in beliefs about teaching and learning mathematics.

Textbooks. The appendix provides information about the textbooks used by schools during each year of the study and the number of students on which complete data (i.e., prior achievement data and scores on two end-of-year tests administered each year) were available. For the purpose of this study, schools using NSF-funded curricula are grouped together for subsequent analyses, a decision consistent with that of the NRC:

> [W]e report the patterns of results separately for evaluations of NSF-supported and commercially generated programs because NSF-supported programs had a common set of design

specifications consistent with the National Council of Teachers of Mathematics (NCTM) *Standards*, reliance on manipulatives, drawing topics from statistics, geometry, algebra and functions, and discrete mathematics at each grade level, and strong use of calculators and computers. (2004, pp. 7–8)

Although the NSF-funded curricula share design features, they are distinctive from one another, having been written by different author teams. We concede that, in “lumping together” textbook series into one of two curriculum types (NSF funded or publisher developed), we sacrifice direct comparisons of one textbook to another. Nevertheless, similar groupings of textbooks by curriculum type are used in recent debates of “new ‘new math’” and “traditional” school mathematics programs (Kilpatrick, 2003; Senk & Thompson, 2003).

Documenting the Implementation of District-Adopted Textbooks

Documenting the extent to which textbooks influence the content of and activity within mathematics lessons (treatment integrity) is a necessary component of any curriculum evaluation (NRC, 2004). In our study, measuring treatment integrity required a standard that could be applied uniformly to teachers using different textbooks, including NSF-funded and publisher-developed textbooks. To establish the degree to which students’ classroom experiences were influenced by the district-adopted textbook, we set a minimum standard regarding textbook use, including (1) regular use of the curricular materials by teacher and students over a sustained period of time, (2) use of a significant portion of the curricular materials for content focus and instructional design, and (3) utilization of instructional strategies aligned with the philosophical embodiment of the textbook. To this end, a number of instruments were developed or adapted to document treatment integrity, including Textbook-Use Diaries and Observation Protocols (each used in both years), and a Table-of-Contents Implementation Record (used in Year 2 only).

Textbook-Use Diary. During three 10-day intervals (in October, January, and March), teachers recorded the curricular materials used in planning and enacting mathematics instruction. They specifically noted what materials were used in (a) planning instruction, (b) implementing instruction, and (c) assigning homework. Additionally, teachers noted whether students utilized their textbooks during each lesson and their utilization of additional curricular resources (e.g., alternative textbooks or other supplemental material). From the 30-day sample, inferences were made regarding the frequency of textbook use by teachers and students for the entire academic year.

Observation Protocol. We adapted an observation tool used in the Wisconsin Longitudinal Study (Romberg & Shaffer, 2003) to measure the classroom learning environment (a process described in a subsequent section) and to document teacher and student interactions with textbooks and other curricular materials. In particular, we recorded (a) textbook use by *teachers* and *students* during the observed lesson, (b) homework assigned from the textbook or other resources, (c) the degree to which the textbook influenced the *content* of the lesson (using a 5-point rubric), and

(d) the degree to which the textbook influenced the *presentation* of the lesson (using a 5-point rubric). Observations were scheduled to occur on days in which teachers made entries in Textbook-Use Diaries, thus corroborating these self-reported entries.

Table-of-Contents Implementation Record.[2] Consistent with the call to document the "extent of coverage" of curricular materials (NRC, 2004), we sought to identify the set of textbook lessons that students experienced. Because teachers are generally given autonomy over what and how to teach, we documented which mathematics content strands were given relative emphasis when teachers enacted the textbook curriculum. In particular, we photocopied the table of contents of textbooks, and teachers highlighted the lessons (or investigations) they taught. These records were collected at four times during the academic year and provided insight into what content from the textbook students had an opportunity to learn.

Validity of teachers' self-reported data. Although multiple sources of data were used to document the implementation of curricular materials, we relied heavily on teachers' self-reported data. Porter (2002) discussed the question of validity of teacher survey data. He stated that "a number of investigations of the validity of survey data for reporting instructional practice have been completed, especially in the context of mathematics instruction" (p. 9). Porter indicated that teacher surveys are valid only when teachers are willing to complete them. In this study, completion of Textbook-Use Diaries and Table-of-Contents Implementation Records were required, and all teachers complied with requests for data on extent of coverage. Data from these sources can be judged to be valid, given that "survey data is excellent for describing . . . what gets taught and for how long" (p. 9). However, teachers must validly describe their practices and are more likely to do so when data are not used for teacher evaluation purposes. In this study, participating teachers were promised that results would be aggregated by *curriculum type*, not reported at the teacher level, even when results were shared with district administrators in participating schools. This assurance of anonymity of research findings promoted the validity of self-reported data because teachers were cognizant that they were not being directly evaluated. (For a robust review of literature on tools for measuring the content of instruction, see Porter, 2002.)

Quantifying Implementation: Development of an Implementation Index

To operationally quantify implementation of district-adopted textbooks, we utilized a variety of data sources to create a numerical index that was used in subsequent quantitative analysis of student achievement data. Specifically, a *composite implementation index* was developed, comprising the following components: (1) relative frequency of textbook use by teacher during instruction, (2) relative frequency of textbook use by students during instruction, (3) relative frequency of textbook use for homework assignment, (4) influence of textbook on lesson content,

[2] We acknowledge that these data were collected in Year 2 only, following an Advisory Board recommendation.

(5) influence of textbook on lesson presentation, and (6) textbook coverage. Data for the first three components were based on Textbook-Use Diaries and represent the *percentage of instructional days* the textbook was used by the teacher, students, and as a source of homework, respectively. The fourth and fifth variables were based on observational data and represent an average (over three observed lessons) of the perceived impact of the textbook on lesson content and presentation using a 4-point scale (0 = *not at all,* 1 = *very little,* 2 = *somewhat,* 3 = *a great deal*). Finally, based on data from Table-of-Contents Implementation Records, the sixth variable represents the *percentage* of available textbook lessons or investigations that were reported as "taught."

It was our consensus that the six factors should not be weighted equally in formulating a composite implementation index. For example, we argue that the extent-of-textbook-coverage data are more important in conveying student experiences with the textbook curriculum than other factors such as how frequently students use their textbook. Accordingly, four members of the research team independently assigned weights to all six factors, and these judgments were aggregated to yield relative weightings for each variable comprising the composite implementation index (see Table 1).[3]

Table 1
Relative Weights of the Variables Used to Calculate the Composite Implementation Index

Variable	Weight (year 1)	Weight (year 2)
Relative frequency of textbook use by teacher during instruction (percentage of instructional days)[a]	28%	21%
Relative frequency of textbook use by students during instruction (percentage of instructional days)[a]	10%	8%
Relative frequency of textbook use for homework assignment (percentage of instructional days)[a]	14%	11%
Influence of textbook on lesson content[b]	27%	21%
Influence of textbook on lesson presentation[b]	21%	17%
Textbook coverage (percentage of lessons/investigations taught)[c]	N/A	21%

Note. Totals may not sum to 100 due to rounding.
[a]As reported in the Textbook-Use-Diary.
[b]As recorded in the Observation Protocol.
[c]As reported in the Table-of-Contents Implementation Record.

Although there was consensus on which elements should comprise the implementation index, we concede that the relative weighting assigned to each factor is ultimately subjective. On the other hand, opportunity to learn—defined by the NRC

[3] We note that because the extent-of-textbook-coverage data were not collected in Year 1, there is a precipitous drop in weights of the five other variables in Year 2. The weighting of the other five variables, relative to one another, remains in a constant proportion across both years.

as “circumstances that allow students to engage in and spend time on academic tasks” (2001a, p. 333)—is “widely considered the single most important predictor of student achievement” (p. 334). Thus, characterization of the enacted curriculum must include documentation of the *source* and *implementation* of mathematical tasks (Stein, Remillard, & Smith, 2007), including activities used in classroom instruction and problems assigned as homework. Concomitant with these notions, the composite implementation index includes the relative frequency of textbook use by teachers and students in order to identify the source(s) and implementation of mathematics tasks that comprise students’ opportunity to learn.

Measuring the Standards-Based Learning Environment

The implementation index provided a measure of the extent to which students experienced regular use of curricular materials as well as their opportunity-to-learn mathematics from district-adopted textbooks. Nevertheless, it did not capture the degree to which teaching practices were aligned with curricular materials. We concede that it is difficult to identify an “instructional philosophy” associated with every textbook used in this study. However, NSF-funded curricular materials were purposefully developed to embody tenets of the NCTM *Standards*, and the materials reflect the curriculum developers’ interpretation of these standards. It is our position that a “faithful” implementation of an NSF-funded curriculum ostensibly requires that teachers establish and maintain a learning environment consistent with the NCTM *Standards*. However, this expectation does not uniformly apply across all textbooks in our sample. On the contrary, some textbook publishers (e.g., Saxon) disavow many tenets of the *Standards*. Because many publisher-developed textbooks do not explicitly stipulate the use of standards-based instructional practices, observation of such practices would not ostensibly represent evidence of a “faithful” implementation of such textbooks. Nevertheless, use of publisher-developed textbooks certainly does not preclude teachers from utilizing instructional practices aligned with the *Standards*. Therefore, we documented the degree to which *all* students (regardless of curriculum type) learned within an environment consistent with the recommendations of NCTM, and this necessitated classroom observations.

Observation Protocol. As previously noted, we adapted an observation tool used in the Wisconsin Longitudinal Study (Romberg & Shaffer, 2003) to document the extent to which students experienced a standards-based learning environment. Operationally, observers utilized a rubric[4] to indicate the degree (on a scale of 1 to 3) to which each of the following five classroom events were evident during each observed lesson:

1. The enacted lesson provided opportunities for students to make conjectures about mathematical ideas.

[4] Space limitations prohibit inclusion of these rubrics. However, they are available upon request from the authors.

2. The enacted lesson fostered the development of conceptual understanding.
3. Students explained their responses or solution strategies.
4. Multiple perspectives/strategies were encouraged and valued.
5. The teacher valued students' statements about mathematics and used them to build discussion or work toward shared understanding for the class.

The research team was trained in the use of the Observation Protocol and application of its accompanying rubric. Training consisted of collaborative discussions of the protocol, and several rounds of simulated observations using videotaped lessons. Through negotiation, agreement was reached on 89% of all codings.

For each of the five classroom events, we determined an overall code (High, Medium, or Low) based on data from three observations. In particular, we summed individual scores (1, 2, or 3) to determine whether each event should be characterized as High (7 or greater), Medium (5 to 6), or Low (4 or less). Categorical codes corresponded to numerical codes as follows: High = 2, Medium = 1, and Low = 0. We then summed the five numerical codes to yield a composite score ranging from 0 to 10, an interval scale. In this regard, 0 essentially represents an absence of a standards-based learning environment. The composite score was used to classify teachers' standards-based learning environment into three categories: High (7 or greater), Medium Moderate (3 to 6), and Low (2 or less). It should be noted that these classifications represent 1 year of student experiences. To classify student experiences over 2 years, we define the Standards-Based Learning Environment (SBLE) as the sum of the composite scores in Year 1 and Year 2. The SBLE code was used as an independent variable in subsequent analyses of student achievement data.

Student Achievement Measures

The NRC (2004) identified numerous problems in the selection and use of outcome measures. Most notably, "too many tests . . . rely exclusively on multiple-choice format, limiting the assessment of the cognitive levels of performance and neglecting the long-term development of student knowledge" (p. 195). The NRC recommended the use of multiple-outcome measures that vary by question type (open ended, multiple choice) and type of test (international, national, local), and that ensure "curricular validity of measures," meaning the outcome measures are "sensitive to the curriculum's stated goals and objectives" (p. 49). Baseline and outcome measures of student achievement used in this study reflected many of these recommendations but were not designed to be particularly sensitive to one curriculum type (NSF funded or publisher developed) nor to any single textbook per se.

Baseline measures. Prior achievement data were gathered from existing school records and used as a covariate in analyses. In most cases, schools provided a standardized mathematics achievement test score (Iowa Test of Basic Skills [ITBS] or Stanford Achievement Test 9 [SAT9]) collected during the previous spring for each

student. These are nationally normed exams with sound psychometric properties (Geisinger, Spies, Carlson, & Plake, 2007; Spies & Plake, 2005). For two suburban schools in the same state, the only available measure was from an assessment instrument developed by CTB-McGraw Hill for that particular state; this assessment instrument was consistent in test format and largely similar in content focus with tests designed for use on a national level. All baseline data were converted to Normal Curve Equivalents (NCE) in order to provide a standard format for subsequent comparative purposes.

TerraNova Survey. Published by CTB-McGraw Hill, the TerraNova Survey (TNS) is a norm-referenced, multiple-choice test that exhibits sound psychometric properties (Spies & Plake, 2005), and it provided a common measure across all sites. This outcome measure professes to measure student performance in all five Content Standards (Number and Operations, Algebra, Geometry, Measurement, and Data Analysis and Probability) and to focus on both mathematical skills and concepts. A grade-specific TNS was administered in the spring of each year to all Grade 6–8 students in the sample. Forty minutes were allowed for the test. Calculators were not permitted on the first eight items on each exam but were allowed thereafter, as per instructions from the publisher. Scale scores were used in quantitative analyses because they use the same metric across grade levels.

Balanced Assessment in Mathematics. At the time of data collection, the Balanced Assessment in Mathematics (BAM) was published by CTB-McGraw Hill (but is now published by Mathematics Assessment Resource Service). In contrast to TNS, the BAM is a criterion-referenced test. The BAM is a grade-specific, constructed-response test that professes to measure mathematical reasoning, problem solving, and communication, as well as proficiency in skills and procedures (Spies & Plake, 2005). A grade-specific test was administered in the spring of each year to all Grade 6–8 students in the sample. Each test contained five contextualized problems, within which students answered short fill-in-the-blank questions and longer show-your-work or explain-your-reasoning questions. All responses were scored using a rubric, and the test had a maximum of 40 points. Student performance is more broadly gauged by the assignment of one of four performance levels (see Table 2).

Following administration of the student assessments in the spring of 2003 and 2004, a senior staff member of the BAM project trained the research team and approximately two dozen local middle and secondary mathematics teachers to score student responses to the BAM. These scorers did not know the identity of the schools as they scored tests. In an independent recoding of a stratified random sample of 171 tests, the average total score differed by less than 0.2 points (on a 40-point scale), indicating a high degree of reliability of scores.

Consistent with recommendations for the design of comparative studies of mathematics curricula (NRC, 2004), the tests used in this study were of the highest quality—nationally recognized, assessing important mathematics, and comprising items that were refined after large-scale pilot testing. Additionally, our choice of

Table 2
Performance Levels on the Balanced Assessment in Mathematics

Level of performance	Description
Level 1	The student's response shows few of the elements of performance that the tasks demand.
Level 2	The student's response shows some of the elements of performance that the tasks demand and some signs of a coherent attack on the core of some of the problems. However, the shortcomings are substantial, and the evidence suggests that the student would not be able to produce high-quality solutions without significant further instruction.
Level 3	For most tasks, the student's response shows the main elements of performance that the tasks demand, organized as a coherent attack on the core of the problems. There are errors or omissions, some of which may be important, but of a kind that the student could well fix, with more time for checking and revision and some limited help.
Level 4	The student's response meets the demands of nearly all of the tasks, with few errors. With some more time for checking and revision, excellent solutions would seem likely.

tests ensured that outcome measures would vary by question type (multiple choice and constructed response) and type of test (norm referenced and criterion referenced). However, given the fundamentally different purposes and properties of these exams, it would not be psychometrically sound to "combine" the two outcome measures into one "hybrid" dependent variable. Such a composite score could not be meaningfully interpreted in ways that scale scores (TNS) and performance levels (BAM) can be interpreted separately.

ANALYSIS OF DATA

In this study, the substantive focus was to examine how variables at the teacher level such as the use of one curriculum type and/or presence of a standards-based learning environment impact middle school students' mathematics achievement. In studies of teacher-level effects, the data are inherently nested. Students are nested within classrooms, classrooms are nested within teachers, teachers are nested within schools, and so on. Similar to Harwell et al. (2007), HLM methods were used to explicitly address the nesting of the data (for methodological details, see Bryk & Raudenbush, 1992). There are three major issues with the analysis of multilevel data using single-level methods such as ANOVA or OLS regression: (a) underestimation of standard errors, (b) aggregation bias, and (c) heterogeneity of regression slopes.

The first issue is that of *standard errors*. A curriculum is experienced by students in a group, and therefore individual student responses and what they learn are ostensibly correlated. Stated alternatively, a student's test score is not independent of his or her other test scores and are more correlated with each other than to other

students' scores. Similarly, students from the same classroom share more characteristics than they do with students from other classrooms. Because of this dependence, the estimation of the standard errors used for statistical testing are overestimated in a single-level analysis (e.g., ANOVA) at the student level, which would result in an underestimate of the statistical significance of the group-level effects. Curricular evaluation studies have recently been criticized for inappropriate selection of unit of analysis (Kilpatrick, 2003; NRC, 2004). In most studies, *student* is the unit of analysis, essentially (and incorrectly) assuming that each student had an independent experience with the curriculum. By employing HLM, we address this critical issue by analyzing nested data.

A second issue is that of *aggregation bias*. A variable can have a different meaning or effect at different levels. For example, the gender of a student could influence his or her mathematics achievement score. But the gender composition of the student's classroom or school could have a different effect. Multilevel modeling allows for the study of the effects of variables at the different levels.

The third issue concerns the *heterogeneity of regression slopes*. Here, student level variables (e.g., gender, ethnicity, SES) and mathematics achievement could vary across schools and be influenced by school-level variables. A common example in mathematics education is the study of achievement gaps. One could easily imagine a case in which achievement gaps were different from school to school, and the research study attempted to attribute these gaps to specific school characteristics. Multilevel modeling allows for the formal exploration of whether relationships between student-level variables and outcome variables vary between schools, and if so, what school-level variables are important.

Finally, similar to Hill, Rowan, and Ball (2005) and many studies involving large-scale assessments, there was missing data on a number of variables. As in the Hill analyses, mean imputation was used to deal with missing cases; however in this sample, the amount of missing data on the variables used in the analyses was small. The largest amount of missing data on any individual variable was 1.9%. For a full description of the mean imputation process, we direct readers to Little and Rubin (1987). For all HLM analyses conducted in this study, the final sample included 2533 students, 33 teachers at the time of assessment, and 10 schools.

RESEARCH FINDINGS

Drawing on the data collected over 2 years, this section summarizes the main findings of the study. It is organized around the primary research questions.

Predicting Student Achievement From Implementation Indices

In Table 3, we report the mean composite implementation indices for teachers of each curriculum type, NSF funded and publisher developed. Additionally, we report means for each of the six variables comprising the composite implementation index. For 4 of 6 variables, scores are *percentages*, thus on a scale from 0 to

Table 3
Textbook Implementation Data for Teachers in Years 1 and 2

	Year 1			Year 2		
Variable and curriculum type	*N*	Mean	SD	*N*	Mean	SD
Composite implementation index						
NSF funded	24	78.77	11.308	16	77.71	12.140
Publisher developed	24	73.29	15.714	23	76.23	7.497
Relative frequency of textbook use by teacher during instruction (percentage of instructional days)						
NSF funded	24	80.67	12.802	16	87.38	11.916
Publisher developed	24	76.13	15.540	23	81.13	13.509
Relative frequency of textbook use by students during instruction (percentage of instructional days)						
NSF funded	24	77.79*	12.914	16	83.44*	11.183
Publisher developed	24	66.46*	21.514	23	70.74*	21.009
Relative frequency of textbook use for homework assignment (percentage of instructional days)						
NSF funded	24	58.54	20.347	16	73.00	18.195
Publisher developed	24	56.75	26.995	23	68.17	22.340
Influence of textbook on lesson content[a]						
NSF funded	24	2.6528	.43383	16	2.5625	.39849
Publisher developed	24	2.5556	.55313	23	2.6377	.41333
Influence of textbook on lesson presentation[a]						
NSF funded	24	2.3333	.58772	16	2.2083	.61914
Publisher developed	24	2.0556	.64206	23	2.3406	.59983
Textbook coverage (percentage of lessons/investigations taught)						
NSF funded	24	N/A	N/A	16	67.63	21.341
Publisher developed	24			23	68.13	16.168

[a]Four-point scale: 0 = *not at all*, 1 = *very little*, 2 = *somewhat*, 3 = *a great deal.*
$^*p < .05$.

100. With regard to the perceived influence of the district-adopted textbook on lesson content and presentation, scores range from 0 (*not at all*) to 3 (*a great deal*).

As reported in Table 3, in both Year 1 and Year 2 the mean composite implementation indices were not significantly different across groups of teachers, with each group scoring in the mid-to-upper 70s, reflecting relatively strong implementation. That is, there were no significant differences with regard to implementation between teachers using NSF-funded or publisher-developed textbooks—both groups of teachers utilized their textbooks at approximately the same level.

Because the implementation index is a linear combination of many factors, we undertook a secondary analysis of data on each variable comprising the index. Specifically, we detected no significant differences between the two groups of teachers on any singular variable in Year 1 or 2, with one exception. Teachers using

NSF-funded curricula reported a significantly higher Frequency of Use of Textbooks by Students During Instruction, and this higher rate was observed in both Year 1 (77.79%, NSF vs. 66.46%, publisher developed) and Year 2 (83.44%, NSF vs. 70.74%, publisher developed).

Data on teachers' extent of coverage of curricular materials in Year 2 are reported elsewhere (see Tarr et al., 2006) but revealed that both sets of teachers placed greater emphasis on Number and Operations than would be expected, given the composition of their textbooks, often at the expense of Data Analysis and Probability, Geometry, and Measurement. It is interesting to note that all teachers of *Connected Mathematics* and *Mathematics in Context* utilized at least some modules that were intended for students in a different grade and in nearly all cases the modules were written for lower grades, but this phenomenon largely reflected strict district policies regarding scope and sequence.

We utilized the implementation indices to examine whether implementation is a predictor of student achievement; that is, we sought to determine whether there is a relationship between implementation indices and student achievement on the BAM and TNS. HLM revealed that the composite implementation index was not a significant predictor of mathematics achievement, using the BAM performance levels and TNS scale scores as separate dependent variables. Stated alternatively, higher indices were not necessarily associated with higher student achievement on either outcome measure, and this finding is discussed in a later section.

Although Table 3 includes means and standard deviations, it does not reveal that one teacher reported not using her district-adopted textbook, *MathThematics*. This virtual abandonment of the textbook curriculum yielded her a composite implementation index of 7 on a 100-point scale (more than five standard deviations below the mean of teachers using NSF-funded curricula). Although there was some variability in implementation indices, we ultimately excluded only this teacher from analysis of student achievement data. Because our index takes into account different facets of implementation, for all but this aforementioned teacher, we considered textbook use to be substantial enough to examine links between curriculum type and student outcome measures.

Predicting Student Achievement From Curriculum Type

Both the BAM and TNS outcomes were analyzed separately with HLM. Student-level regression models containing prior mathematics achievement, gender, and ethnicity were fitted to the data for each classroom. For each outcome (BAM and TNS), three models were fitted. First, an unconditional model of the form

$$Y_{ijk} = \pi_{0jk} + e_{ijk}$$
$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$
$$\beta_{00k} = \gamma_{000} + u_{ook}$$

was fitted, where Y_{ijk} is the achievement of child i in classroom j and school k, π_{0jk} is the mean mathematics achievement of classroom j in school k, β_{00k} is the mean mathematics achievement in school k, γ_{000} is the grand mean, e_{ijk} is the random

student effect, r_{0jk} is a random classroom effect, and u_{ook} is a random school effect. Results of the unconditional models showed that 8% and 15.4% of the variation in mathematics achievement was between teachers and schools for the TNS and BAM, respectively.

To address the effect of curriculum type on student achievement, a model of the form

$$\text{Level-1 Model: } Y_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{Prior Achievement}_{ijk} - \overline{X}_1) + \pi_{2jk}(\text{Gender}_{ijk} - \overline{X}_2) + \pi_{3jk}(\text{Black}_{ijk} - \overline{X}_3) + \pi_{4jk}(\text{Hispanic}_{ijk} - \overline{X}_4) + \pi_{5jk}(\text{Native American}_{ijk} - \overline{X}_5) + \pi_{6jk}(\text{Other Ethnicity}_{ijk} - \overline{X}_6) + r_{ijk}$$

$$\text{Level-2 Model: } \pi_{0jk} = \beta_{00k} + \beta_{01k}\text{NSF}_{jk} + r_{0jk}$$

$$\text{Level-3 Model: } \beta_{00k} = \gamma_{000} + u_{00k}$$

was fitted, where π_{1jk} is the student-level slope capturing the effect of prior mathematics achievement with teacher *j* in school *k*, and β_{01k} is the teacher-level slope capturing the effect of NSF-funded curricula on mathematics achievement in school *k*. The effects of curriculum type (NSF funded or publisher developed) on mathematics achievement, using the BAM performance levels and TNS scale scores as separate dependent variables, are shown in Table 4. In all cases, the main effect of curriculum type was not found to be statistically significant.

Table 4
Effects of Curriculum Type on the BAM and the TNS $n_{Level\ 1} = 2533$; $n_{Level\ 2} = 33$

	BAM	TNS
Mean mathematics achievement	1.955*** (.071)	701.632*** (2.362)
NSF funded	.216 (.122)	–0.812 (4.041)
Prior achievement	.033*** (.001)	1.689*** (.040)
Gender	–.019 (.029)	–3.652** (1.284)
Black	.015 (.070)	–5.055 (3.119)
Hispanic	–.005 (.065)	–5.029 (2.881)
Native American	.011 (.141)	8.742 (6.301)
Other ethnicity	.071 (.071)	5.284 (3.134)

**$p < .01$. ***$p < .001$.

Prevalence and Predictive Power of a Standards-Based Learning Environment (SBLE)

Borko, Kieran, and Lester (2004) note that "the general mathematics education community too often uses terms such as 'standards-based instruction,' 'reform-based classrooms,' 'problem-based instruction,' and 'inquiry-based teaching' interchangeably" (p. 168). For the purpose of this study, SBLE classified student experiences over 2 years with respect to five "classroom events" that we have associated with "standards-based instruction."

Figure 2 depicts, by textbook, the SBLE score (0 to 10) for each teacher in the 2-year study. We observed lower levels of standards-based instruction in each year of the study, yet ranging widely in both years. As shown in the three left-most columns, a standards-based learning environment was more prevalent among teachers utilizing an NSF-funded curriculum. However, it is also clear that although NSF-funded curricula may encourage or facilitate use of these standards-based prac-

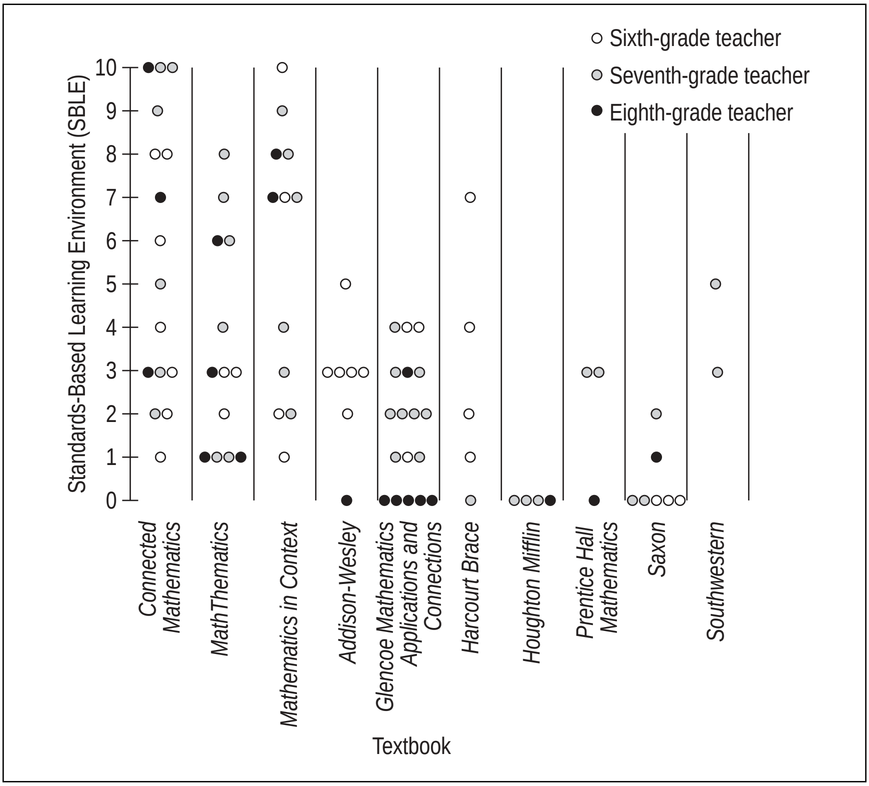

Figure 2. Standards-Based Learning Environment (SBLE), by teachers and textbook.

tices, they are obviously not sufficient in yielding High levels, as evident in relatively Low scores for approximately 30% of all teachers using this curriculum type. Teachers of publisher-developed curricula also exhibited generally Low SBLE scores, but this was not uniformly the case; Moderate levels of SBLE were evident among teachers of Glencoe, Harcourt Brace, and Southwestern.

Interaction Between Curriculum Type and SBLE

To address research question 3, the previous model was changed at Level 2:

$$\text{Level-2 Model: } \pi_{0jk} = \beta_{00k} + \beta_{01k}(\text{NSF})_{jk} + \beta_{02k}(\text{SBLE})_{jk} + \beta_{013k}(\text{NSF} \times \text{SBLE})_{jk} + r_{0jk}$$

This final model included the addition of SBLE as well as the interaction between curriculum type and SBLE.

Table 5 shows the effects of SBLE, curriculum type, and the interaction between the two on mathematics achievement using the BAM and TNS standardized tests as separate dependent variables. Using the BAM standardized test as the dependent variable, the interaction between SBLE and curriculum type was found to be statistically significant. Using the TNS as the dependent variable, the interaction and the main effects of SBLE and curriculum type were not statistically significant.

Table 5
Effects of Curriculum Type and SBLE on BAM and TNS $n_{Level\ 1} = 2533$; $n_{Level\ 2} = 33$

	Balanced Assessment in Mathematics	TerraNova Survey
Mean mathematics achievement	2.041***	704.681***
	(.114)	(4.218)
NSF funded	–0.446	–15.667
	(.251)	(9.311)
Standards-Based Learning Environment (SBLE)	–0.023	–0.796
	(.025)	(.910)
NSF × SBLE	0.080*	1.989
	(.032)	(1.170)
Prior achievement	.033***	1.687***
	(.001)	(.040)
Gender	–.020	–3.684**
	(.029)	(1.284)
Black	.018	–5.035
	(.070)	(3.125)
Hispanic	–.001	–5.054
	(.065)	(2.889)
Native American	.004	8.600
	(.141)	(6.302)
Other ethnicity	.077	5.365
	(.071)	(3.143)

*$p < .05$. **$p < .01$. ***$p < .001$.

The joint relationship between curriculum type and SBLE on BAM performance level is shown in Figure 3. To aid in interpretation, we trichotomize the SBLE variable based on our previous definition to Low, Moderate, and High to depict the trend in student performance level on the BAM; most notably, the slope is both significant and positive. That is, students in classrooms were positively impacted on the BAM by NSF-funded curricula when coupled with either Moderate or High levels of SBLE. There was no statistically significant impact of NSF-funded curricula on students in classrooms with a Low level of SBLE. By way of contrast, although there appears to be a slightly negative relationship between publisher-developed textbooks and SBLE, our analysis revealed that this relationship is not statistically significant. Stated differently, students in classrooms with publisher-developed textbooks were not impacted on the BAM by varying levels of SBLE. It should also be pointed out that there was no significant impact of either curriculum type when coupled with varying levels of SBLE on the TNS as the dependent measure.

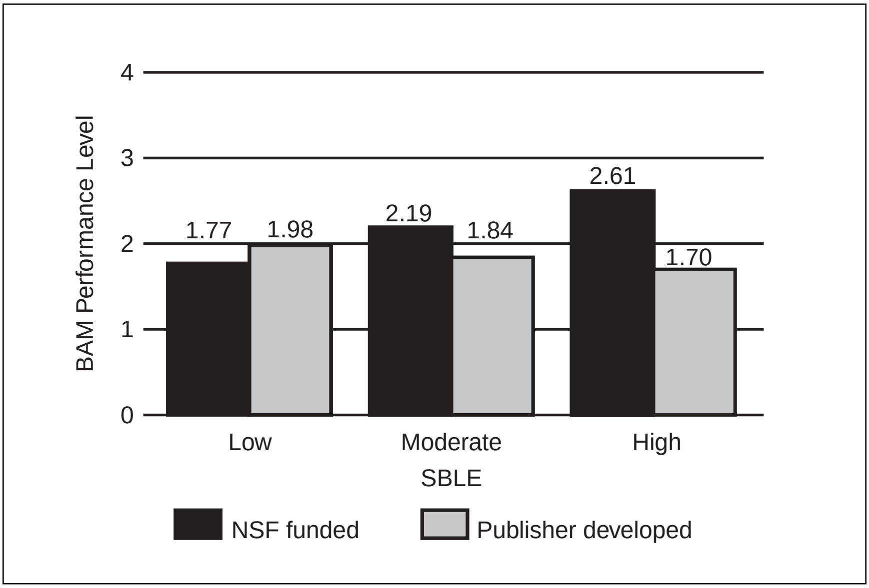

Figure 3. Effect of SBLE on BAM performance level, by curriculum type.

LIMITATIONS

We acknowledge several limitations of this study. These include issues related to the sample and analyses, limited observation data, and challenges associated with equating groups, as well as school, teacher, and student attrition.

Sample Size at the Teacher Level

Through the use of HLM, we were able to statistically address the nesting of the data, analyzing the data as students within teachers within schools. However, because the final sample size was 33 teachers at the time of the assessment, this profoundly limited the statistical modeling that we were able to do. With this small sample size, it was difficult to add more variables at the teacher level because of the limited statistical power. From a more general perspective, there are numerous research questions that are of interest at the teacher/classroom/school level. With multilevel modeling techniques, the sample sizes at each of these levels need to be sufficient to address the research questions. Issues of statistical underspecification can be addressed with larger sample sizes.

One Urban School Was Dropped From the Study

This decision was made because of numerous problems associated with data collection. Initially it was difficult to obtain prior achievement scores of the sixth graders and seventh graders, and the school district promised but ultimately refused to provide FRL data at the student level. The withdrawal of one teacher, responsible for more than half of all seventh graders, coupled with high student absenteeism and incomplete prior achievement data resulted in exclusion of this school from analysis. Notwithstanding the loss of statistical power resulting from attrition, two other urban schools remained in our sample for the duration of the study.

One School Changed Its Mathematics Program After the 1st Year

No mention of the possibility of changing mathematics curriculum arose during the 1st year of the study. Yet during the summer between Year 1 and 2, the teachers at one school chose to return to a textbook series formerly used. Although this decision was not unanimous among teachers, all agreed to abide by this decision. Consequently, this school, originally identified as a user of an NSF-funded curriculum, used the program for 1 year and then switched to a publisher-developed curriculum (see Appendix, school 5).

Direct Comparisons of Students Participating in Different Mathematics Programs Was Not Possible

Three different NSF mathematics programs were represented in this study. However, companion schools (similar in demographics to schools using NSF-developed textbooks) used more than 20 different publisher-developed mathematics textbooks, and these programs differed in content focus and instructional philosophy. As a result of the diversity of textbooks used, it was not appropriate to directly compare the achievement of students experiencing one textbook against students experiencing another textbook. Instead, student achievement data were analyzed with respect to curriculum type (NSF funded vs. publisher developed). The decision to group all three NSF-funded textbooks is consistent with similar

groupings of previous efforts to reform school mathematics; most notably, a variety of curriculum projects were "lumped together" in the 1960s to form a corpus of curricula known as "the new math" (Davis, 1990).

No Common Prior Achievement Measures Across All Schools Were Available

This study relied on available prior student performance data rather than asking schools to commit additional days to testing.[5] To document prior student achievement, six schools provided ITBS scores, two SAT9, and the two remaining schools provided results on their state assessments. It is worth noting that the two remaining schools were a matched pair: one school used an NSF-funded textbook, one used a publisher-developed textbook; they were located in the same state; and they provided achievement data on their (common) state assessment. In an effort to get a common metric for prior achievement, a transformation was done to produce NCEs. This transformation to a common metric was necessary but results on the same standardized test would have been ideal for a prior achievement measure.

Limited Number of Teacher Observations

Each teacher was observed three times during each year of the study. Although teachers may have taught several classes of the same type, the researchers did not observe every class. Lack of resources prevented more observations. More observations during the year, including observations in every class taught by every teacher, would likely have strengthened the profile on the classroom learning environment. Given this limitation, it is not possible to generalize the results from the observational data to all of the teachers' classes. Consequently this reduced the number of classes, thereby sacrificing some statistical power in subsequent analyses. Nevertheless, data were generally consistent across three observations both in relation to the classroom learning environment and to teachers' use of curricular materials. Stated alternatively, teachers who had a low SBLE in a single observation were typically assigned identical codes in *each* of the three independent observations. Similar consistency was evident among teachers with higher SBLEs.

DISCUSSION

In the wake of No Child Left Behind (NCLB, 2002), school districts are required to monitor achievement trends and document "Adequate Yearly Progress." Declines in performance, stagnant test scores, or even modest improvements that do not meet public expectations often become an impetus for changes in curriculum policy, beginning with the selection of district-adopted textbooks. Results of this study

[5] Schools were unwilling to commit to more than 2 days of project-related testing in addition to existing state testing programs. Consequently, it was not possible to administer a common prior achievement measure in our study.

underscore the complexities associated with curriculum evaluation research and caution against oversimplistic interpretations of student achievement data in relation to mathematics curriculum. In fact, the quasi-experimental design of this study implies that strong causal inferences are not warranted because students were not randomly assigned to classes, classes were not randomly assigned to teachers, and teachers were not randomly assigned to one of two curriculum types. Moreover, consistent with our conceptual framework (see Figure 1), we concede that numerous factors other than curriculum affect student performance in mathematics. Nevertheless, there are several key messages that emerge from the research findings reported here.

Documenting Treatment Integrity

The discordant use of district-adopted textbooks by middle-grades mathematics teachers in this study likely reflects the autonomy of many U.S. teachers who operate under the tenets of “local control.” Thus, teacher decisions regarding the selection of mathematics content to teach and emphasize resulted in striking differences in students’ opportunity to learn, even within a given school building (Tarr et al., 2006). This result is consistent with Kilpatrick’s (2003) assertion that “two classrooms in which the same curriculum is supposedly being ‘implemented’ may look very different” (p. 473). Furthermore, these disparities in students’ opportunity to learn suggest the need for greater oversight by supervisors of mathematics education at all levels.

From a research perspective, if links are to be established between student outcomes and mathematics curriculum, these gaping discrepancies in teachers’ textbook use necessitate the documentation of treatment integrity, or “fidelity of implementation.” Results of our study suggest that this, too, is a complex endeavor. Indeed, we acknowledge methodological challenges associated with measuring what we refer to as *textbook integrity* (for a full description, see Chval et al., in press). Our implementation indices included measures of *how often* teachers used curricular materials and the *extent of coverage* of textbooks. We observed markedly different enactments of the same lesson in which textbooks were “used” and lessons “covered,” resulting in distinctive experiences as students interacted with curricular materials in classrooms. Moreover, when teachers deviated from textbooks, they supplemented in markedly different ways; several teachers of NSF-funded curricula created worksheets to strengthen students’ procedural skills, whereas at least some teachers using publisher-developed textbooks supplemented with investigations from *Connected Mathematics Project*. Our composite implementation index was not sensitive enough to capture these nuances, and perhaps this might explain why implementation fidelity was not singularly a significant predictor of student achievement. Future research should seek to develop a metric for detecting student achievement in relation to particular uses of curricular materials, to identify essential factors of implementation, and to determine appropriate weightings for each factor.

Results of the BAM

After taking into account student background variables (most notably prior achievement) and teacher variables (e.g., implementation indices), curriculum type was not a significant predictor of student achievement on either student outcome measure. However, our models ascertained a significant relationship between curriculum type, the classroom learning environment, and student performance on the BAM. In particular, in classrooms where teachers offered Moderate or High levels of SBLE with an NSF-funded curriculum, there were significantly higher performance levels on the BAM (see Figure 3). The fact that such an interaction was evident for only NSF-funded curricula is noteworthy for several reasons. First, NSF-funded programs had a common set of design specifications stipulating that each textbook embody major tenets of the NCTM *Standards.* In fact, many NSF-funded textbooks were initially (and continue to be) marketed as "*Standards*-based," offering an alternative to more traditional school mathematics programs. Thus, it is clearly expected that any implementation of a "*Standards*-based" curriculum include the provision of a standards-based learning environment. When teachers of an NSF-funded curriculum enacted the curriculum in a manner consistent with authors' intent, the results were superior performance on constructed-response items comprising the BAM, a format increasingly used in many state-mandated testing programs. On the other hand, when teachers taught an NSF-funded curriculum *without* a standards-based learning environment, incongruence between curriculum and instruction resulted and was manifested in lower achievement on the BAM. Stated alternatively, an NSF-funded curriculum without a standards-based learning environment can no longer be considered an implemented "standards-based" curriculum, and such implementation was associated with a generally lower performance level on the BAM.

A similar argument can be employed to explain why an interaction between SBLE and publisher-developed textbooks did not yield the same achievement trends on the BAM. In particular, because publisher-developed textbooks do not stipulate employment of "standards-based instruction," it should not be surprising that SBLE was not a significant predictor of student achievement in classrooms with this curriculum type. In short, the neither higher nor lower levels of SBLE necessarily represent incongruence between curriculum and instruction when publisher-developed textbooks are being used.

It should be noted that higher levels of SBLE were more prevalent among teachers utilizing NSF-funded curricula. Stated differently, a standards-based learning environment was more readily created by these teachers. However, the availability of "*Standards*-based" curricular materials did not translate into High (or even Moderate) levels of SBLE for all teachers using NSF-funded textbooks. This finding suggests that ongoing professional development is needed because, without the provision of professional development aimed at yielding a more faithful implementation of NSF-funded curricula, student performance may be expected to lag.

The fact that a low level of SBLE was evident among 30% of teachers using NSF-funded curricula is troubling given the composition of the school and teacher samples. After all, Principal Investigators of NSF IMD grants *nominated* these schools, teachers in these schools received significantly more professional development than teachers of publisher-developed textbooks, and professional development generally focused on learning how to faithfully implement the NSF-funded textbook. Such variable enactment of NSF-funded curricula by teachers in this study is further evidence that "standards-based curricula are challenging to enact well" (Stein et al., 2007, p. 361). Additional studies are needed to identify the factors that influence successful implementation of standards-based curricula and explore why curriculum-specific professional development fails to achieve desired goals for some participants.

Results of the TNS

The lack of significant findings in TNS data is compelling for several reasons. One possible explanation for the lack of significant findings relates to the selection of outcome measures used in curriculum evaluation studies. As Kilpatrick (2003) pointed out, "Standardized achievement tests, in particular, are exceedingly blunt instruments for measuring what students might learn in a given year from a given curriculum" (p. 479), because they contain only a few items per content strand, mathematical process or concept. In fact, the NRC (2004) stipulated the need to establish "curricular validity of measures" (p. 49) when drawing inferences between student achievement and mathematics curriculum. We did not seek to determine the curricular validity of the TNS or BAM and therefore cannot support the position that either outcome measure has (or lacks) curricular validity. Such claims would have necessitated detailed curriculum analyses of more than 20 mathematics textbooks in relation to 32 multiple-choice test items, and this was beyond the scope of our resources. However, given that the TNS, like the BAM, is considered to be a respectable instrument for gauging student achievement in mathematics, it is possible that both curriculum types yielded comparable student performance on the TNS. Such a finding would be consistent with previous research (e.g., Senk & Thompson, 2003) that indicated that students who experienced NSF-funded curricula perform at least comparable to students who experienced more traditional school mathematics programs.

A second plausible explanation for the lack of significant differences is that, after accounting for initial differences in student achievement, there was simply little variation in TNS scores. As previously stated, for all participating schools, prior achievement was provided in the form of NCEs on nationally normed, multiple-choice tests. These standardized achievement tests are highly correlated with one another. In one study (Virginia Department of Education, 1997), a .81 correlation was found between scores on the ITBS and SAT9 tests of eighth-grade mathematics. The ITBS, SAT9, and TNS purport to measure the same general construct, and all three instruments share the same stated purpose, namely *to compare student performance* to a norming

population. Thus, after adjusting for initial differences in prior achievement, scores on the TNS may have offered little variation that could be attributable to one curriculum type or another, yielding a lack of statistically significant differences.

Selecting Outcome Measures

Our decision to use the TNS as an outcome measure was largely based on the fact that local decision makers and the general public place great value on student performance within the school system or in relation to national norms. When seeking schools to participate, we found that school district personnel embraced the idea that we would shed light on how their students performed compared to national norms. In this regard, at least some school administrators seemed most concerned about whether their students were progressing at the same rate as their peers in other schools or whether they were performing above or below average relative to a national sample. Stated alternatively, superintendents and principals at some schools were less concerned about what mathematics students precisely learned but instead were preoccupied about “how they did” overall. This message is consistent with Kilpatrick (2003) who suggested that recent curriculum evaluation research is aimed at reassuring the public that NSF-funded curricula are not harming students rather than assessing what students are actually learning. Thus, the limited utility of scores on the TNS calls into question the use of norm-referenced achievement tests in curriculum evaluation studies: “Norm-referenced data are limited, however, because they do not show what a student actually can or cannot do. . . . Nor does the norm-referenced score indicate what a student needs to do to improve” (NRC, 2001b, pp. 213–214).

In contrast, because the BAM is a criterion-referenced test, it enabled the reporting of each student’s performance in absolute terms—that is, in terms of what the student *can or cannot do*. Performance assessments have the added value of enabling administrators (and researchers) to examine student achievement across student populations, including students who experienced a particular curriculum type (NSF funded or publisher developed). However, curriculum evaluation studies are in need of performance assessments that precisely measure what is learned from a given curriculum or, in comparative studies, what is learned from two or more curricula. Future research should seek to develop curriculum-specific or “fair” tests that exhibit curricular validity of outcome measures and provide content-specific subscores that can convey the relative strengths and weaknesses of a curriculum to textbook authors and publishers as well as to parents, teachers, administrators, and policymakers.

CONCLUSION

This study sought to compare and contrast the systemic effects of textbook adoptions, curriculum, teacher practices, and standards-based learning environment on student achievement in mathematics. The findings reported herein answer some research questions but raise several more. Given that students using both types of

curriculum performed comparably on both the TNS and BAM, why does there continue to be an ongoing controversy surrounding the NSF-funded curricula? This quasi-experimental study joins an expanding body of research that heretofore has revealed essentially no significant differences in student achievement when considering only curriculum type; achievement was significantly impacted when considering the interaction between curriculum type and SBLE.

Additionally, this study provides evidence that a standards-based learning environment positively impacts achievement on performance assessments such as the BAM but only when coupled with a curriculum that is designed to embody this pedagogical orientation. In particular, when NSF-funded curricula were implemented in a manner consistent with its design specifications, student achievement on the BAM was compelling; when this was not the case, BAM performance levels were lower. It follows that what is needed is *coherence* between the textbook and implemented curricula; that is, consistency between curriculum and instruction is needed in order to actualize student learning in mathematics.

The importance (yet overt lack) of a standards-based learning environment raises the following questions: Why does professional development successfully impact some teachers but not others? What makes implementation of a NSF-funded curriculum so difficult that it precipitates glaring deviations from the textbook or, in at least one case, the abandonment of an entire school mathematics program? Future research should address these questions if curriculum continues to be viewed by policymakers as a primary lever for improving student achievement.

Finally, this study addressed the need for additional research into curricular effectiveness (NRC, 2004) but nonetheless represents only one piece. Additional research is needed to ascertain the relationship between mathematics curricula and student learning, taking into account teachers' interpretation of, interaction with, and implementation of district-adopted textbooks and other curricular materials.

REFERENCES

Bay-Williams, J. M., Reys, B. J., Reys, R. E. (2003). Effectively implementing standards-based mathematics curricula in middle schools. *Middle School Journal, 34* (4), 36-41.

Ben-Peretz, M. (1990). *The teacher-curriculum encounter: Freeing teachers from the tyranny of texts*. Albany: State University of New York Press.

Borko, H., Kieran, C., & Lester, F. K. (2004). Teaching and learning. In F. K. Lester & J. Ferrini-Mundy (Eds.), *Proceedings of the NCTM research catalyst conference* (pp. 163–169). Reston, VA: National Council of Teachers of Mathematics.

Briars, D. (March, 2001). *Mathematics performance in the Pittsburgh public schools*. Paper presented at a conference of the Mathematics Assessment Resource Service, San Diego, CA.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.

Chval, K., Chávez, Ó., Reys, B. J., & Tarr, J. E. (in press). Considerations and limitations related to conceptualizing and measuring textbook integrity. In J. T. Remillard, B. Herbel-Eisenmann, & G. Lloyd (Eds.), *Teachers' use of mathematics curriculum materials: Research perspectives on relationships between teachers and curriculum*.

Chval, K., Reys, R. E., Reys, B. J., Tarr, J. E., & Chávez, Ó. (2006). Pressures to improve student performance: A context that both urges and impedes school-based research. *Journal for Research in Mathematics Education, 37*, 158–166.

Clandinin, D. J., & Connelly, F. M. (1992). Teacher as curriculum maker. In P. W. Jackson (Ed.), *Handbook of research in curriculum: A project of the American Educational Research Association* (pp. 369–401). New York: Macmillan.

Clarke, B., Clarke, D., & Sullivan, P. (1996). The mathematics teacher and curriculum development. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 1207–1233). Dordrecht, The Netherlands: Kluwer.

Davis, R. B. (1990). Discovery learning and constructivism. In R. B. Davis, C. A. Maher, & N. Noddings (Eds.), *Constructivist views on the teaching and learning of mathematics* (*JRME* Monograph 4, pp. 93–106). Reston, VA: National Council of Teachers of Mathematics.

Geisinger, K. F., Spies, R. A., Carlson, J. F., & Plake, B. S. (Eds.). (2007). *The seventeenth mental measurements yearbook.* Lincoln, NE: Buros Institute of Mental Measurements.

Griffin, L. A., Evans, A., Timms, T., & Trowell, J. (2000). *Arkansas grade 8 benchmark exam: How do Connected Mathematics schools compare to state data?* Little Rock: Arkansas State Department of Education.

Grouws, D. A., & Smith, M. S. (2000). NAEP findings on the preparation and practices of mathematics teachers. In E. A. Silver & P. A. Kennedy (Eds.), *Results from the seventh mathematics assessment of the National Assessment of Educational Progress* (pp. 107–139). Reston, VA: National Council of Teachers of Mathematics.

Harwell, M. R., Post, T. R., Maeda, Y., Davis, J. D., Cutler, A. L., Andersen, E., et al. (2007). Standards-based mathematics curricula and secondary students' performance on standardized achievement tests. *Journal for Research in Mathematics Education, 38*, 71–101.

Hiebert, J. (1999). Relationships between research and the NCTM Standards, *Journal for Research in Mathematics Education, 30*, 3–19.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*, 371–406.

Huntley, M. A., Rasmussen, C. L., Villarubi, R. S., Sangtong, J., & Fey, J. T. (2000). Effects of standards-based mathematics education: A study of the *Core-Plus* mathematics project algebra and functions strand. *Journal for Research in Mathematics Education, 31*, 328–361.

Jackson, P. W. (Ed.). (1992). *Handbook of research on curriculum: A project of the American Educational Research Association*. New York: Macmillan.

Jennings, J. & Rentner, D. (2006). Ten big effects of the No Child Left Behind Act on public schools. *Phi Delta Kappan, 88*, 110–113.

Kilpatrick, J. (2003). What works? In S. L. Senk & D. R. Thompson (Eds.), *Standards' based school mathematics curricula. What are they? What do students learn?* (pp. 471–493). Mahwah, NJ: Lawrence Erlbaum.

Lappan, G., Fey, J., Fitzgerald, W., Friel, S., & Phillips, E. D. (1998). *Connected Mathematics Project*. Palo Alto, CA: Dale Seymour.

Lindquist, M. M. (1997). NAEP findings regarding the preparation and classroom practices of mathematics teachers. In P. Kenney & E. A. Silver (Eds.), *Results from the Sixth Mathematics Assessment of the National Assessment of Educational Progress*. Reston, VA: National Council of Teacher of Mathematics.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.

Love, E., & Pimm, D. (1996). "This is so": A text on texts. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 371–410). Dordrecht, The Netherlands: Kluwer.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, et al. (2000). *TIMSS 1999 international mathematics report: Findings from IEA's repeat of the Third International Mathematics and Science Study at the eighth grade.* Chestnut Hill, MA: International Study Center, Boston College.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Research Council. (2001a). *Investigating the influence of standards: A framework for research in mathematics, science, and technology education.* Washington, DC: The National Academies Press.

National Research Council. (2001b). *Knowing what students know: The science and design of educational assessment*. Washington, DC: The National Academies Press.

National Research Council. (2004). *On evaluating curricular effectiveness: Judging the quality of K–12 mathematics evaluations*. Washington, DC: The National Academies Press.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Peak, L. (1996). *Pursuing excellence: A study of U.S. eighth-grade mathematics and science teaching, learning, curriculum, and achievement in international context. Initial findings from the Third International Mathematics and Science Study* (No. NCES-97-198). Washington, DC: Office of Educational Research and Improvement.

Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31*(7), 3–14.

Porter, A. C., & Smithson, J. L. (2001, December). Defining, developing, and using curriculum indicators (CPRE Research Report No. RR-048). Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education. Retrieved February 3, 2003, from http://www.cpre.org/images/stories/cpre_pdfs/rr48.pdf.

Remillard, J. T. (1999). Curriculum materials in mathematics education reform: A framework for examining teachers' curriculum development. *Curriculum Inquiry, 29,* 315–342.

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, *75*, 211–246.

Remillard, J. T., & Bryans, M. B. (2004). Teachers' orientations toward curriculum materials: Implications for curricular change. *Journal for Research in Mathematics Education, 35,* 352–388.

Reys, B. J. (Ed.) (2006). *The intended mathematics curriculum as represented in state-level curriculum standards: Consensus or confusion?* Greenwich, CT: Information Age Publishing, Inc.

Reys, B. J. & Reys, R. E. (2006). The development and publication of elementary mathematics textbooks: Let the buyer beware! *Phi Delta Kappan*, 87, 377–383.

Reys, R. E., Reys, B. J., Lapan, R., Holliday, G., & Wasman, D. G. (2003). Assessing the impact of standards-based middle grades mathematics curriculum materials on student achievement. *Journal for Research in Mathematics Education, 34,* 74–95.

Reys, B. J., Robinson, E., Sconiers, S., & Mark, J. (1999). Mathematics curricula based on rigorous national standards: What, why, how? *Phi Delta Kappan, 80*, 454–456.

Riordan, J. E., & Noyce, P. E. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education, 32,* 368–398.

Robitaille, D. F., Schmidt, W. H., Raizen, S., McKnight, C., Britton, E., & Nicol, C. (1992). *TIMSS monograph 1: Curriculum frameworks for mathematics and science*. Vancouver, Canada: Pacific Educational Press, University of British Columbia.

Romberg, T. A. & Shafer, M. C. (2003). Mathematics in Context (MiC)—Preliminary evidence about student outcomes. In S. Senk & D. Thompson (Eds.) *Standards-based school mathematics curricula: What are they? What do students learn?* (pp. 225–284). Mahwah, NJ: Lawrence Erlbaum Associates.

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., et al. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics, 102,* 245–253.

Schmidt, W. H., McKnight, C. C., & Raizen, S. A. (1997). *A splintered vision: An investigation of U. S. science and mathematics education*. Hingham, MA: Kluwer.

Schoenfeld, A. H. (2000). Purposes and methods of research in mathematics education. *Notices of the American Mathematical Society, 47,* 641–649.

Senk, S. and Thompson, D. (Eds.). (2003). *Standards-based school mathematics curricula: What are they? What do students learn?* Mahwah, NJ: Lawrence Erlbaum Associates.

Sewall, G. T. (2005). Textbook publishing. *Phi Delta Kappan, 86,* 498–502.

Spies, R. A., & Plake, B. S. (Eds.). (2005). *The sixteenth mental measurements yearbook.* Lincoln, NE: Buros Institute of Mental Measurements.

Stein, M. K., Remillard, J., & Smith, M. S. (2007). How curriculum influences student learning. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 319–370). Charlotte, NC: Information Age Publishing.

Tarr, J. E., Chávez, Ó., Reys, R. E., & Reys, B. J. (2006). From the written to the enacted curricula: The intermediary role of middle school mathematics teachers in shaping students' opportunity to learn. *School Science and Mathematics, 106,* 191–201.

Thomas B. Fordham Institute. (2004). *The mad, mad world of textbook adoption.* Washington, DC: Author.

Trafton, P. R., Reys, B. J., & Wasman, D. G. (2001). Standards-based mathematics curriculum materials: A phrase in search of a definition. *Phi Delta Kappan, 83,* 259–264.

Travers, K. J., & Westbury, I. (1989). *The IEA study of mathematics I: Analysis of mathematics curricula.* Oxford: Pergamon Press.

Tyson, H. (1997). *Overcoming structural barriers to good textbooks.* Paper commissioned by the National Education Goals Panel. Retrieved October 11, 2005, from http://govinfo.library.unt.edu/negp/Reports/tyson.htm

Tyson-Bernstein, H., & Woodward, A. (1991). Nineteenth century policies for twenty-first century practice: The textbook reform dilemma. In P. G. Altbach, G. P. Kelly, H. G. Petrie & L. Weis (Eds.), *Textbooks in American society* (pp. 91–104). Albany: State University of New York Press.

Usiskin, Z. (1998). Which curriculum is best? *UCSMP Newsletter, 24,* 3–10.

Virginia Department of Education. (1997). The Virginia State Assessment Program: Transition from the Iowa Tests of Basic Skills and the tests of academic proficiency to the Stanford 9. Richmond: Author.

Weiss, I. R., Banilower, E., McMahon, K., & Smith, P. S. (2001). *Report of the 2000 National survey of science and mathematics education*. Chapel Hill, NC: Horizon Research.

Authors

James E. Tarr, 303 Townsend Hall, Department of Learning, Teaching and Curriculum, University of Missouri, Columbia, MO 65211-2400; tarrj@missouri.edu

Robert E. Reys, 303 Townsend Hall, Department of Learning, Teaching and Curriculum, University of Missouri, Columbia, MO 65211-2400; reysr@missouri.edu

Barbara J. Reys, 303 Townsend Hall, Department of Learning, Teaching and Curriculum, University of Missouri, Columbia, MO 65211-2400; reysb@missouri.edu

Óscar Chávez, 303 Townsend Hall, Department of Learning, Teaching and Curriculum, University of Missouri, Columbia, MO 65211-2400; chavezo@missouri.edu

Jeffrey Shih, Department of Curriculum and Instruction, 4505 Maryland Parkway, University of Nevada—Las Vegas, Las Vegas, NV 89154-3005; jshih@unlv.nevada.edu

Steven J. Osterlind, 16 Hill Hall, Department of Educational, School and Counseling Psychology, University of Missouri, Columbia, MO 65211-2130; osterlinds@missouri.edu

APPENDIX

Textbooks used by school, grade, and course (Year 1):

School	Grade	Course	Textbook	No. of students
2	6	6B	*Connected Mathematics*	53
	7	7B	*Connected Mathematics*	47
3	6	6B	*MathThematics*	86
	7	7B	*MathThematics*	58
		7C	*MathThematics*	19
4	6	6B	*MathThematics*	195
		6C	*MathThematics*	23
	7	7B	*MathThematics*	206
		7C	*MathThematics*	27
5	6	6A	*Mathematics in Context*	33
		6B	*Mathematics in Context*	110
	7	7B	*Mathematics in Context*	132
6	6	6B	*Mathematics in Context*	138
		6C	*Mathematics in Context*	40
	7	7B	*Mathematics in Context*	127
		7C	*Mathematics in Context*	31
7	6	6A	*Harcourt Brace*	59
		6B	*Harcourt Brace*	181
		6C	*Harcourt Brace*	24
	7	7A	*Glencoe Mathematics Applications and Connections*	50
		7B	*Glencoe Mathematics Applications and Connections*	111
		7C	*Glencoe Mathematics Applications and Connections*	172
8	6	6B	*Addison-Wesley*	105
	7	7B	*Southwestern*	79
		7C	*Southwestern*	17
9	6	6B	*Saxon*	41
	7	7A	*Houghton Mifflin*	45
10	6	6B	*Glencoe Mathematics Applications and Connections*	108
		6C	*Glencoe Mathematics Applications and Connections*	34
	7	7B	*Glencoe Mathematics Applications and Connections*	56
		7C	*Glencoe Mathematics Applications and Connections*	68
11	6	6B	*Saxon*	49
		6C	*Saxon*	19
		6A	*Saxon*	16
	7	7B	*Saxon*	35
		7C	*Saxon*	18
		7A	*Saxon*	7

Note: The digit in the course code refers to the grade, the letters designate the level of the class, as reported by the school: A, remedial or basic; B, regular; C, advanced.

Textbooks used by school, grade, and course (Year 2):

School	Grade	Course	Textbook	No. of students
2	7	7B	*Connected Mathematics*	53
	8	8B	*Connected Mathematics*	47
3	7	7B	*MathThematics*	61
		7C	*MathThematics*	25
	8	8B	*MathThematics*	51
		8C	*Integrated Mathematics: Houghton Mifflin*	26
4	7	7B	*MathThematics*	19
		7C	*MathThematics*	26
	8	8B	*MathThematics*	205
		8C	*Mathematics: Modeling Our World Southwestern*	28
5	7	7A	*Glencoe Mathematics Applications and Connections*	30
		7B	*Glencoe Mathematics Applications and Connections*	92
		AA	*Focus On Algebra SF-AW*	21
	8	8A	*Glencoe Mathematics Applications and Connections*	8
		8B	*Glencoe Mathematics Applications and Connections*	13
		8C	*Glencoe Mathematics Applications and Connections*	88
		AA	*Focus On Algebra SF-AW*	23
6	7	7B	*Mathematics in Context*	135
		7C	*Mathematics in Context*	43
	8	8B	*Mathematics in Context*	125
		AA	*Algebra I McDougal Littell*	33
7	7	7A	*Glencoe Mathematics Applications and Connections*	65
		7B	*Glencoe Mathematics Applications and Connections*	74
		7C	*Glencoe Mathematics Applications and Connections*	103
		7C	*UCSMP Transition Mathematics*	22
	8	8A	*Addison-Wesley Middle School Math*	61
		8B	*Addison-Wesley Middle School Math*	14
		8C	*Glencoe Mathematics Applications and Connections*	169
		AA	*Algebra 1 Glencoe*	89
8	7	7B	*Prentice Hall Mathematics*	77
		7C	*Prentice Hall Mathematics*	28
	8	8B	*Prentice Hall Mathematics*	78
		8C	*Algebra 1 Holt Rinehart Winston*	18
9	7	7B	*Houghton Mifflin Mathematics Experience*	41
	8	8B	*Mathematical Connections Houghton Mifflin*	45
10	7	7B	*Glencoe Mathematics Applications and Connections*	117
		7C	*Pre-Algebra Glencoe*	25
	8	8B	*Glencoe Mathematics Applications and Connections*	21
		8C	*Pre-Algebra Glencoe*	80
		AA	*Algebra I McDougal Littell*	23
11	7	7A	*Saxon Math 87*	18
		7B	*Saxon Math 87*	41
		7C	*Saxon Algebra 1/2*	25
	8	8A	*Saxon Algebra 1/2*	5
		8B	*Saxon Algebra 1/2*	31
		8C	*Saxon Algebra 1*	24

Note: The digit in the course code refers to the grade, the letters designate the level of the class, as reported by the school: A, remedial or basic; B, regular; C, advanced. AA represents an algebra course.